

Introduction to Groundwork: Formula of Universal Law (xvii–xxiii)

Christine M. Korsgaard

October 7. Introduction, Second section through The Formula of Universal Law, book pages xvii–xxiii.

Source: *Groundwork of the Metaphysics of Morals*, Cambridge University Press, revised edition, 2012. Translation by Mary Gregor, revised by Jens Timmermann; introduction by Christine M. Korsgaard.

This text and its PDF are converted from the supplied ebook. PDF page numbers are not the printed book page numbers. Kant's Akademie references are retained. The wording, emphasis, and notes come from the book.

Second section

Although the argument of the first section proceeded from our ordinary ideas about morality, and involved the consideration of examples, it is not therefore an empirical argument. The examples do not serve as a kind of data from which conclusions about moral motivation are inductively drawn. Instead, the argument is based on our rational appraisal of the people in the examples, taking the facts about their motivation as given: if these people act from respect for law, as the examples stipulate, then their actions have moral worth. Whether anyone has ever actually acted from respect for law is a question about which moral philosophy must remain silent. So demonstrating that the categorical imperative governs our wills is not a matter of showing that we actually act on it. Instead, it is a matter of showing that we act on it in so far as we are rational. A comparison will help here. Showing that the principle of non-contradiction governs our beliefs is not a matter of showing that no one ever in fact holds contradictory beliefs, for people surely do. Nor is it a matter of showing that people are sometimes moved, say, to give up cherished beliefs when they realize those beliefs will embroil them in contradiction. Instead, it is a matter of showing that in so far as they are rational, that's what they do. Kant's project in the second section therefore is to:

trace and distinctly present the practical rational faculty from its general rules of determination up to where there arises from it the concept of duty. (4:412)

In other words, in the second section Kant lays out a theory of practical reason, in which the moral law appears as one of the principles of practical reason.

It is a law of nature, very roughly speaking, that what goes up must come down. Toss this book into the air, and it will obey that law. But it will not, when it reaches its highest point, say to itself, "I ought to go back down now, for gravity requires it." As rational beings, however, we do in this way reflect on, and sometimes even announce to ourselves, the principles on which we act. In Kant's words, we act not merely in accordance with laws, but in accordance with our representations or conceptions of laws (4:412).

Yet we human beings are not perfectly rational, since our desires, fears, and weaknesses may tempt us to act in irrational ways. This opens up the possibility of a gap between the principles upon which we actually act – our maxims or subjective principles – and the objective laws of practical

reason. For this reason, we conceive the objective laws of practical reason as imperatives, telling us what we *ought* to do. The theory of practical reason is therefore a theory of imperatives.

Imperatives may be either hypothetical or categorical. A hypothetical imperative tells you that if you will something, you ought also to will something else: for example, if you will to be healthy, then you ought to exercise. That is an imperative of skill, telling you how to achieve some particular end. Kant believes that there are also hypothetical imperatives of prudence, suggesting what we must do given that we all will to be happy. A categorical imperative, by contrast, simply tells us what we ought to do, not on condition that we will something else, but unconditionally.

Kant asks how all these imperatives are “possible” (4:417), that is, how we can establish that they are legitimate requirements of reason, binding on the rational will. He thinks that in the case of hypothetical imperatives the answer is easy. A hypothetical imperative is based on the principle that whoever wills an end, in so far as he is rational, also wills the means to that end. This principle is analytic, since *willing* an end, as opposed to merely wanting it or wishing for it or thinking it would be nice if it were so, is setting yourself to bring it about, to cause it. And setting yourself to cause something just is setting yourself to use the means to it. Since willing the means is conceptually contained in willing the end, if you will an end and yet fail to will the means to that end, you are guilty of a kind of practical contradiction.

Since a categorical imperative is unconditional, however, there is no condition given, like the prior willing of an end, which we can simply analyze to derive the “ought” statement. The categorical imperative must therefore be *synthetic*, so morality depends on the possibility of establishing a *synthetic a priori* practical principle.

The Formula of Universal Law

Kant does not, however, move immediately to that task; in fact, he will not be in a position to take that up until the third section. The second section, like the first, proceeds “analytically.” Kant is still working towards uncovering *what* we have to prove *in order to* establish that moral requirements really bind our wills. The first step is to analyze the very idea of a categorical imperative in order to see what it “contains.” Kant says:

when I think of a categorical imperative I know at once what it contains. For since besides the law the imperative contains only the necessity of the maxim to conform with this law, whereas the law contains no condition to which it was limited, nothing is left but the universality of a law as such, with which the maxim of the action ought to conform, and it is this conformity alone that the imperative actually represents as necessary. (4:420–421)

This is the sort of thing that makes even practiced readers of Kant gnash their teeth. A rough translation might go like this: the categorical imperative is a law, to which our maxims must conform. But the reason they must do so cannot be that there is some *further* condition they must meet, or some *other* law to which they must conform. For instance, suppose someone proposed that we must keep our promises because it is the will of God that we should do so – the law would then “contain the condition” that our maxims should conform to the will of God. This would yield only a conditional requirement to keep our promises – “if you would obey the will of God, then you must keep your promises” – whereas the categorical imperative must give us an *unconditional*

requirement. Since if the imperative is to be categorical there can be no such condition, all that remains is that the categorical imperative should tell us that our maxims themselves must be laws – that is, that they must be universal, that being the characteristic of laws.

There is a simpler way to make this point. What could make it true that we must keep our promises because it is the will of God? That would be true only if it were true that we must indeed obey the will of God, that is, if “obey the will of God” were itself a *categorical* imperative. Conditional requirements give rise to a regress; if there are unconditional requirements, we must at some point arrive at principles on which we are required to act, not because we are commanded to do so by some yet higher law, but because they are laws in themselves. The categorical imperative, in the most general sense, tells us to act on *those* principles, principles which are laws in themselves. Kant continues:

There is, therefore, only a single categorical imperative and it is this: *act only according to that maxim through which you can at the same time will that it become a universal law.* (4:421)

Kant next shows us how this principle serves to identify our duties, by showing us that there are maxims which it rules out – maxims which we could not possibly will to become universal laws. He suggests that the way to test whether you can will your maxim as a universal law is by performing a kind of thought experiment, namely, asking whether you could will your maxim to be a law of nature in a world of which you yourself were going to be a part. He illustrates this with four examples, the clearest of which is the second.

A person in financial difficulties is considering “borrowing” money on the strength of a false promise. He needs money, and knows he will get it only if he says to another person, “I promise you I will pay you back next week.” He also knows perfectly well that he will not be able to repay the money when next week comes. His question is whether he can will that the maxim of making a false promise in order to get some money should become a law of nature. Although Kant does not do this, it helps to set out the test in a series of steps.

The first step is to formulate the maxim. In most cases, the person is considering doing a certain act for a certain end, so the basic form of the maxim is “I will do Act-A in order to achieve Purpose-P.” Suppose then that your maxim is:

I will make a false promise in order to get some ready cash.

Next we formulate the corresponding “law of nature.” It would be:

Everyone who needs some ready cash makes a false promise.

At least where duties to others are concerned, Kant’s test may be regarded as a formalization of the familiar moral challenge: “What if everybody did that?” In order to answer this question, you are to imagine a world where everybody does indeed do that. We might call this the “World of the Universalized Maxim.” At this point it is important to notice that Kant says the categorical imperative tells you to act on a maxim which you can *at the same time* will to be a universal law: he means at the same time as you will the maxim itself. So you are to imagine that you are in the World of the Universalized Maxim, seeing whether you can will to act on your maxim in that world. For instance, you imagine that you are asking whether you could will to secure some ready cash by

means of a false promise in a world where everyone who needs a little ready cash (tries to) secure it by means of a false promise. In particular, you are asking whether any contradiction arises when you try to do that. Kant, says, in the example at hand, that it does, because:

the universality of a law that everyone, once he believes himself to be in need, could promise whatever he fancies with the intention not to keep it, would make the promise and the end one may pursue with it itself impossible, as no one would believe he was being promised anything but would laugh about any such utterance, as a vain pretense. (4:422)

Why is this a contradiction? This question has attracted an enormous amount of philosophical attention and many interpretations have been proposed. The views that have been suggested may be divided into three broad categories.

Proponents of a logical contradiction interpretation think Kant means there is a straightforward logical contraction in the proposed law of nature. One might argue, for instance, that the universalization of the maxim of false promising would undercut the very practice of making and accepting promises, thus making promises impossible and the maxim literally inconceivable.⁸

Kant's use of teleological language in some of the examples has suggested to proponents of the teleological contradiction interpretation that the contradiction emerges only when the maxim is conceived as a possible teleological law of nature. False promising violates the "natural purpose" of promising, which is to create trust and cooperation, so that a universal law of false promising could not serve as part of a teleological system of natural laws.

According to proponents of the practical contradiction interpretation, the maxim's efficacy in achieving its purpose would be undercut by its universalization. In willing its universalization, therefore, the agent would be guilty of the same sort of practical contradiction that is involved in the violation of a hypothetical imperative. In fact, the maxim in the example is derived from a hypothetical imperative – "if you need some ready cash, you ought to make a false promise" – which in turn is derived from a "law of nature" or "causal law" – namely that false promising is a cause of, and so a means to, the possession of ready cash. In the World of the Universalized Maxim, however, this law no longer obtains. So in willing the World of the Universalized Maxim the agent undercuts the causal law behind the hypothetical imperative from which his own maxim is derived, making his own method of getting the money ineffective. Language supporting all three of these interpretations can be found in Kant's texts, and different interpretations fit different examples better. The problem of finding a single account of the contradiction test that produces the right answers in all cases is one on which Kantians are still at work.

The question is complicated by the fact that Kant himself thinks contradictions may arise in two different ways (4:421, 4:424). In some cases, he says, the maxim cannot even be thought as a universal law of nature: the contradiction is in the very conception of the universalized maxim as a law. The example we have been considering is of that kind: there could not *be* a law that everyone who needs money should make false promises, so the maxim fails what is often called "the contradiction in conception test." Maxims which fail this test are in violation of strict or perfect duties, particular actions or omissions we owe to particular people, such as the duty to keep a promise, tell the truth, or respect someone's rights. But there are also maxims which we can conceive as universal laws, but which it would still be contradictory to *will* as laws: these maxims

fail what is often called “the contradiction in the will test.” They violate wide or imperfect duties, such as the duty to help others when they are in need, or to make worthwhile use of your talents.⁹ Here again, there is disagreement about exactly what the contradiction is. Kant suggests that “all sorts of possible purposes” (4:423) would have to go unfulfilled in a world in which we had neglected our abilities and in which we could not count on the help of others when we are in need. Since rationality commits us to willing the means to our ends, we must will a world in which these most general means – our own abilities and the help of others – would be available to us.

These examples are offered simply as a few illustrations to show how the categorical imperative works to establish the moral status of our actions. Generally, if a maxim passes the categorical imperative test, the action is permissible; if it fails, the action is forbidden, and, in that case, the opposite action or omission is required. The maxims in the examples fail the test, showing, for instance, that making a false promise is forbidden, and that a commitment to helping others when they are in need is required. For a more complete account of what Kant thinks morality requires of us, however, the reader must look to the *Metaphysics of Morals*.

The thought experiment we have just considered shows us *how* to determine whether a maxim can be willed as a universal law, not *why* we should will only maxims that can be universal laws. Kant is not claiming that it is irrational to perform immoral actions because it actually embroils us in contradictions. The contradictions emerge only when we attempt to universalize our maxims, and the question why we must will our maxims as if they were to become universal laws remains to be answered. It is to this question Kant turns next.

Notes

8. For the notion of a practice and the logical dependence of actions falling under the practice on the existence of the practice itself, see **John Rawls**, “Two Concepts of Rules,” *Philosophical Review* **64** (January 1955): 3–32; reprinted in *John Rawls: Collected Papers*, Cambridge, MA: Harvard University Press, 1999.

9. In the *Groundwork*, Kant lines up the distinction between the contradiction in conception test and the contradiction in the will test with the traditional distinction between perfect and imperfect duties (described above) at 4:421, and with a less familiar distinction between strict or narrow duties and wide duties at 4:424. This parallel might be taken to suggest that these are just two sets of names for the same distinction, or at any rate that the two distinctions coincide. But in the later *Metaphysics of Morals* Kant describes a category of duties which are characterized as perfect duties and yet which, because they are duties of virtue and all of those are wide, must be wide duties (6:421ff.). Kant explains the distinction between narrow and wide obligation in the *Metaphysics of Morals* at 6:390–394. We have a duty of narrow obligation when we are required to perform a particular action, while we have a duty of wide obligation when we are required to adopt a certain general maxim (e.g. to promote the happiness of others) but have leeway as to how to carry the duty out. This explanation leaves the difference between the two distinctions unclear, and Kant never directly addresses the question how the two distinctions are related. If Kant’s considered view is that these two distinctions do not coincide, we are left uncertain whether the contradiction in conception test is best understood as a test for perfect duties, or as a test for strict duties. These rather intricate issues about the categorization of duties matter to the reader of the *Groundwork* because one of the duties Kant uses as an example here – the duty not to commit suicide in order to avoid misery – is one of those apparently identified in the later work as a perfect duty of wide obligation. This should make us cautious about this example.